Usability Studies in Virtual and Traditional Computer Aided Design Environments for Benchmark 2 (Find and Repair Manipulation)

Dr. Syed Adeel Ahmed, Drexel Dr.

Xavier University of Louisiana, New Orleans, LA 70125.
Assistant Professor of Management at Xavier University of Louisiana

Abstract *- A usability study was used to measure user performance and user preferences for a CAVE™ immersive stereoscopic virtual environment with wand interfaces compared directly with a workstation non-stereoscopic traditional CAD interface with keyboard and mouse. In both the CAVE™ and the adaptable technology environments, crystal eye glasses are used to produce a stereoscopic view. An ascension flock of birds tracking system is used for tracking the user's head and wand pointing device positions in 3D space.*
It is argued that with these immersive technologies, including the use of gestures and hand movements, a more natural interface in immersive virtual environments is possible. Such an interface allows a more rapid and efficient set of actions to recognize geometry, interaction within a spatial environment, the ability to find errors, and navigate through a virtual environment. The wand interface provides a significantly improved means of interaction. This study quantitatively measures the differences in interaction when compared with traditional human computer interfaces.
This paper provides analysis via usability study methods for Find and Repair Manipulation termed as Benchmark 2. During testing, testers are given some time to "play around" with the CAVE™ environment for familiarity before undertaking a specific exercise. The testers are then instructed regarding tasks to be completed, and are asked to work quickly without sacrificing accuracy. The research team timed each task, and recorded activity on evaluation sheets for Find and Repair Manipulation Test. At the completion of the testing scenario involving navigation, the subject/testers were given a survey document and asked to respond by checking boxes to communicate their subjective opinions.

Keywords-Usability Analysis, CAVE™ (Cave Automatic Virtual Environments), Human Computer Interface (HCI), Benchmark, Virtual Reality, Virtual Environments, Competitive Comparison,

I. Introduction

This paper is an extension of the work done by Satter (2005) on Competitive Usability Studies of Virtual Environments for Shipbuilding. The key difference is the use of a new immersive environment called CAVE™. The significance and the detail description of this study is very well explained by Satter (2012) in his recent paper. Here we only present the details of this usability study. The CAVE™ was developed at the University of Illinois at Chicago and provides the illusion of immersion by projecting stereo images on the walls and floor of a room-sized cube. Several users wearing lightweight stereo glasses can enter and walk freely inside the CAVE™. A head tracking system continuously adjusts the stereo projection to the current position of the leading viewer. A CAVE™ and wand system schematic is shown in Figure 1&2.

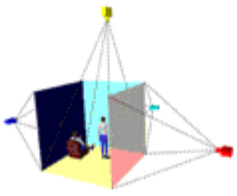


Figure 1: Schematic of the CAVE™ System

Figure 2: The Wand Interface

Environments and Usability Study

The Find and Repair Manipulation scenario was designed to test the user's ability to utilize the two environments/interfaces (Non-stereoscopic workstation and Stereoscopic CAVE™) to Find and Repair Manipulation through the study space locating each of 4 distinct items/parts within the space. The common measure recorded was simply the elapsed time to navigate the space (from a common starting point), locate each

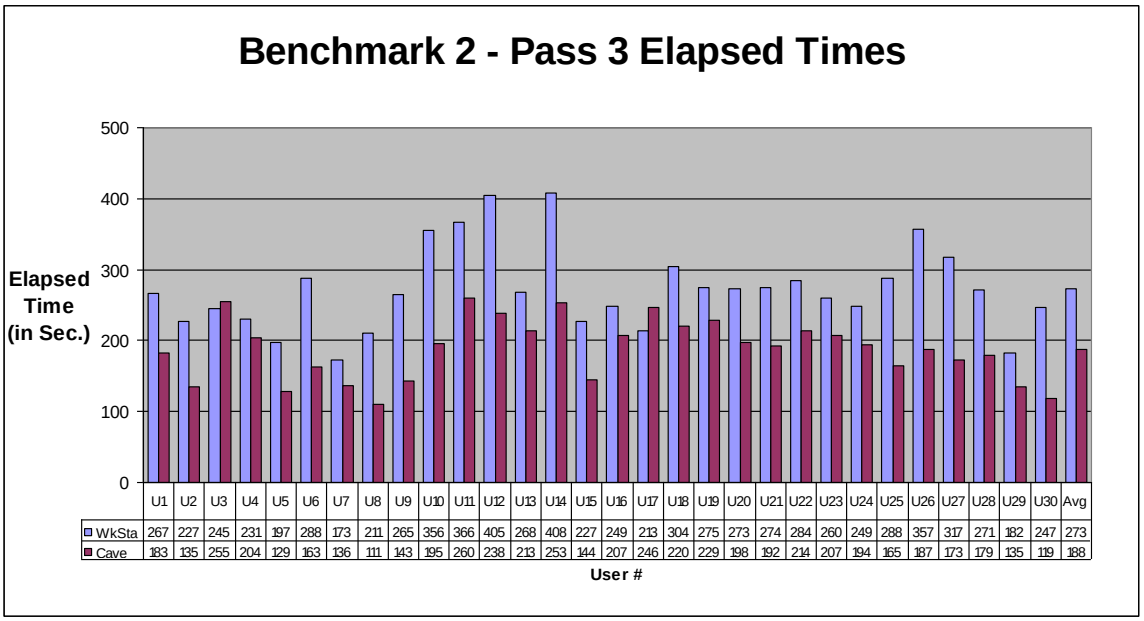
required item/part, and return to the starting point. Each of the thirty users performed this Benchmark three times in each of the two environments. The analysis of the final pass results of these Benchmark 2 tests by the users is presented in the following sections. Pass 3 results represent each user's final exposure to each environment within each scenario. Therefore, pass 3 results tend to show the user's best ability to perform the required tasks. Each environment/interface (Non-stereoscopic workstation and Stereoscopic CAVE™) is represented in a distinct chart.

II. Description

Using the same virtual factory space as used for Benchmark 1[5] , in Benchmark 2 users were required to navigate through the space looking for "errors' that had been injected into the design. Typical "errors" were a screen, turbine or fan, eyewash or conveyor belt, or cyclone separator all placed at a different place from their original place. Users were then required to "fix" the error. The "fix" required the user to utilize the interface (environment) under test (CAVE™, workstation), typically, re-positioning the part to a more suitable location/orientation. Elapsed times were noted for each activity. The elapsed time recorded was the time required to locate and identify the 1st error; the time to "fix" the 1st error; the time to locate and identify the 2nd error; the time to "fix" the 2nd error; and the time to return to the starting position within the space.
The find/repair exercise (Benchmark 2) was repeated three times (three passes) for each of the thirty users in each of the two environments under test and the User Survey was administered to each user after each pass in each environment. As with the Benchmark 1 testing, sequencing of the testers through the two environments was randomized so that not all of the users were testing the same interface at the same time.

Benchmark 2 – pass 3 Elapsed timing analysis:

Figure 1 (Benchmark 2 Pass 3 Elapsed Timings / B2p3Tim) presents a representation of the elapsed times required by users to perform a typical set of find/repair operations as defined in the Benchmark 2 scenarios. The results presented are for the last (3rd) execution of the test. All other results are presented in Appendix B. These times should, and do, represent the "best/fastest" execution times for the group. It should be noted that stereoscopic interface resulted in shorter execution times (as compared to the non-stereoscopic interface). This proves that CAVE™ is faster, efficient and better environment workstation.



	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28	U29	U30	Avg
WkSta	267	227	245	231	197	288	173	211	265	356	366	405	268	408	227	249	213	304	275	273	274	284	260	249	288	357	317	271	182	247	273
Cave	183	135	255	204	129	163	136	111	143	195	260	238	213	253	144	207	246	220	229	198	192	214	207	194	165	187	173	179	135	119	188

Figure 3: B2p3Tim Elapsed Times

Pass-to-Pass Comparison of Elapsed Times Analysis:

B2 Pass to Pass Comparison

	Pass1 to Pass2		Pass2 to Pass 3		Pass1 to Pass3	
	Diff	%	Diff	%	Diff	%
Cave	69.43	21%	66.6	26%	136.03	42%
W/S	79.1	18%	76.9	22%	156	36%

Table1: Pass-to-Pass Comparison of Elapsed Times

Table 1 presents the improvement in find/repair (manipulation) times for users with each successive exposure to each of the two test environments. Note that there appears to beabout a42% improvement in CAVE™ from pass 1 to pass 3 against 36% improvement in workstation from pass 1 to pass 3. This means that the stereoscopic interface appeared to produce reduced find/repair elapsed times over the non-stereoscopic interface.

III. Detailed Statistical Analysis

As described for the Benchmark 1 testing, all statistical analyses of the test data were performed using NCSS. The K.S. normality testing was performed on the Benchmark 2 results. Levene's test was used to test for equal variance of the data. The null hypothesis (H_0) and alternative hypothesis (H_a) as discussed in last paper Benchmark 1 [5] statistical analysis testing applies here (Benchmark 2) as well.

IV. Pass 3 Statistical Analysis

Table 2 presents the descriptive statistics test results normality testing and variance test results of each Benchmark 2 pass 3 dataset by environment. All other results are presented in Appendix B. In this analysis, it is important to note the results of the Kolmogorov-Smirnov test (KS test statistic) for normal (Gaussian) distribution. In this case, note that the pass 3 B2 datasets for the non-stereoscopic environment fail the KS statistic test for normal distribution of the data. Thus the NCSS software performs a nonparametric, Levene's test to test for equal variance.

B2P3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV
Cave	30	187.5	42.7	111	260	>0.10	Yes	22%
W/S	30	272.6	59.3	173	408	<0.10	No	22%
		Homogeneity of Variance			Test for Differences			
		Levene's Test		Equal	Mann-Whitney Test			
		F-Value	P Value	Var?	Value	P Value	Equal?	Significant?
Cave vs W/S		0.94	0.36	Yes	-5.20	<0.001	No	Cave

Table 2: B2p3Tim Elapsed Times Statistics

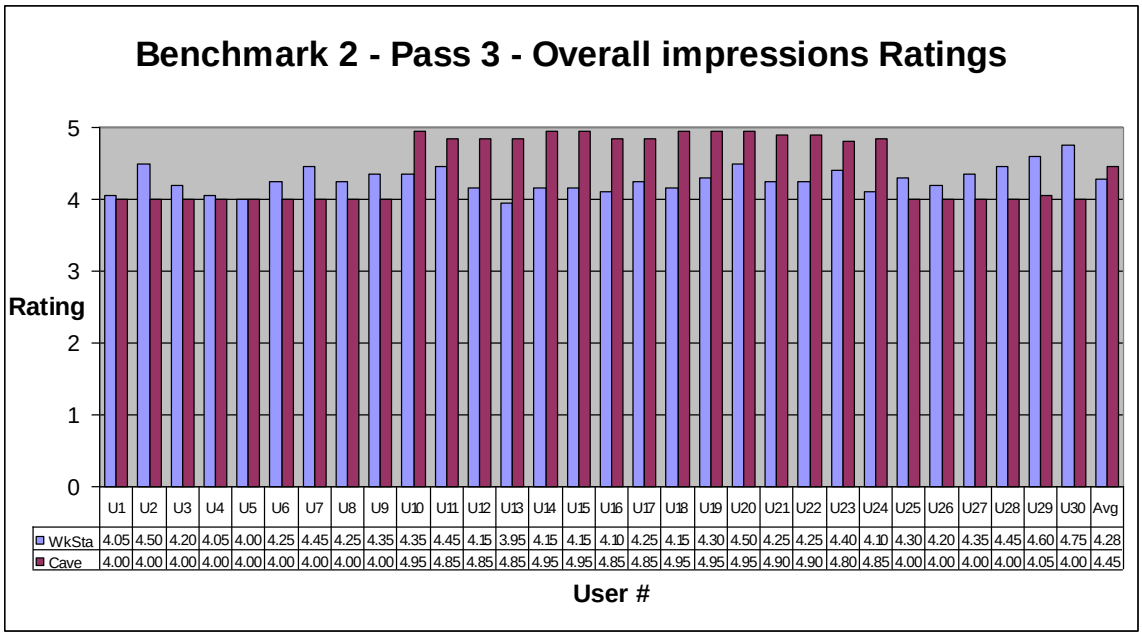
For table 2, the K.S. test is used to test for normality of data. Since the P value is less than 0.1 for workstation, the data are not normal. Levene's test is used to test for equal variance. With that stipulation, since the P value is greater than 0.1, the data have equal variance. Thus, since the data is not normal, Mann Whitney test is used. However, with a Mann Whitney test P value less than 0.1, which indicates that the medians are unequal for the CAVE™ – Workstation comparison. Examination of these results shows that for the two environments, the differences are statistically significant. The conclusion then is that at the 90% confidence level, there is significant evidence to support the alternative hypothesis (H_a). Thus, since the stereoscopic wand environment demonstrates shorter elapsed find/repair times, this environment is statistically "better" than non-stereoscopic workstation environment for Benchmark 2 during pass 3.

V. User Subjective Overall Environment Ratings

After completion of each pass of each Benchmark test in each environment, users provided their subjective views of their experience by completing the 22-question Usability Survey (see Appendix 4) rating the environment on a scale of 1 to 5 (very poor to very good). The questions were grouped into 4 areas (navigation, locating, movement, and general). Following is a presentation of user overall impressions ratings of the interfaces for performing Benchmark 2 tasks (find/repair) at the completion of the 3rd pass as a representation of user final evaluations of each interface. All other results are presented in figure 5.

VI. Benchmark 2 – pass 3 Overall Impressions ratings analysis:

As discussed above, each user was asked to rate his/her experience via the Usability Survey at the completion of each pass of each Benchmark test. Figure 2 (Benchmark 2 pass 3 Overall Impressions ratings /B2p3Ovr) presents the overall impressions ratings of the users at the completion of the 3rd pass of the Benchmark 2 scenario. As such, this represents each user's final impression of the find/repair capabilities of each environment. For Benchmark 2 pass 3 overall impressions ratings, figure 2 shows that user's preferred CAVE™ over workstation.



	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	U22	U23	U24	U25	U26	U27	U28	U29	U30	Avg
WkSta	4.05	4.50	4.20	4.05	4.00	4.25	4.45	4.25	4.35	4.35	4.45	4.15	3.95	4.15	4.15	4.10	4.25	4.15	4.30	4.50	4.25	4.25	4.40	4.10	4.30	4.20	4.35	4.45	4.60	4.75	4.28
Cave	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.95	4.85	4.85	4.85	4.95	4.95	4.85	4.85	4.95	4.95	4.95	4.90	4.90	4.80	4.85	4.00	4.00	4.00	4.00	4.05	4.00	4.45

Figure 2: B2p3Ovr Overall Impressions Ratings

B2OP3	# Users	Mean	St. Dev.	Low	High	P Value	Normal?	CV
Cave	30	4.45	0.45	4	4.95	<0.10	No	10.00%
W/S	30	4.28	0.18	3.95	4.75	>0.10	Yes	4.00%
		Homogeneity of Variance			Test for Differences			
		Levene's Test		Equal	Mann-WhitneyTest			
		F-Value	P Value	Var?	Value	P Value	Equal?	Significant?
Cave vs W/S		176.5	<0.001	No	-0.36	0.36	Yes	N/A

Table3: B2p3Ovr Overall Impressions Ratings

For Table 3, the K.S. test is used to test for normality of data. Since the P value is less than 0.1 for the CAVE™, the data are not normal and Levene's test is used to test for equal variance. Since the P value is less than 0.1 the data have unequal variance. Furthermore, since the data are not normal, theMann Whitney test is used. With aMann Whitney test, P value greater than 0.1, which indicates that the mediansare equal for the CAVE™ and workstation. Examination of these results shows that for the two environments, the differences are not statistically significant. The conclusion then is that at the 90% confidence level, there is significant evidence to support the null hypothesis (H_0). This proves that neither of the two environments is statistically better than each other for Benchmark 2 pass 3 overall impressions subjective ratings.

B2-Pass to Pass Comparison Overall Impressions Ratings Analysis:

Table 4 shows the pass-to-pass improvements in user overall impression ratings for each of the environments. Note that with each successive exposure (pass-to-pass) the user's overall impressions of the interfaces improved. Examination of the pass-to-pass analysis of improvements noted in Table 4 shows that for Benchmark 2, overall impressions subjective ratings, the ratings improved for both CAVE™ and Workstation from pass-to-pass. Comparing CAVE™ and Workstation, the ratings appear to have improved more for CAVE™ with a higher percentage from pass-to-pass than Workstation. Hence, the CAVE™ environment is barely preferred over Workstation for B2 Overall impressions subjective ratings.

B2 Overall Ratings Pass to Pass Comparison

	Pass1 to Pass2		Pass2 to Pass 3		Pass1 to Pass3	
	Diff	%	Diff	%	Diff	%
Cave	-0.7	-20%	-0.18	-4%	-0.88	-25%
W/S	-0.26	-8%	-0.65	-18%	-0.91	-27%

Table 4: B2-Pass to Pass Comparison Overall Impressions Ratings

The negative values in Table 4 show that pass 1 ratings were lower than pass 2 and pass 2 ratings were lower than pass 3. This proves that user's subjective ratings improved from pass to pass. For example a value of -27%

for Workstation (pass 1 to pass 3) is calculated as (3.37-4.28)/3.37, where 3.37 and 4.28 represent the means of Benchmark 2 overall impressions ratings for pass 1 and pass 3 respectively.

Usability Survey

User ID: _____

Environment: 19" CAD Stereo

Pass: 1 2 3

		Very Good 5	Good 4	Neutral 3	Poor 2	Very Poor 1
	Navigation					
1	Initial impression of navigational modes					
2	Gross control movement					
3	Speed of cursor/pointer movement					
4	Ability to make fine adjustments to the placement of the cursor/pointer					
5	Ability to recover cursor/pointer movements					
6	Ease of use					
7	After-test impression of the navigational modes					
	Locating					
1	Initial impression of the interface in locating specific parts/equipment					
2	Ease of identification of selected part/equipment					
3	Ability to make fine adjustments in selecting specific parts/equipment					
4	Ease of use					
5	After-test impression of the location/selection mechanism					
	Movement					
1	Initial impression of the interface for relocating parts/equipment					
2	Ease of movement across the three axis					
3	Ability to make fine part/equipment movement adjustments					
4	Ease of use					
5	After-test impression of the movement mechanism					
	General					
1	Initial impression of the overall system					
2	Ability to relate a 2D planform to the space as presented					
3	"Intuitiveness" of the interface - do the controls follow expected use?					
4	Overall ease of use					
5	After-test impression of the overall system.					

Comments

Figure 5: – Usability Survey Questionnaire (Satter, 2005)

VII. Conclusions

For Benchmark 2 (Find/Repair) the statistics shows better results (lower timings and higher subjective ratings) for the CAVE™ in both objective and subjective measures than the workstation, except for Benchmark 2 pass 3 Location ratings, General, and Overall ratings in which the subjective ratings do not suggest which of the two environments are better. My future work will focus competitive usability on Benchmarks 3 for spatial awareness under these same environments.

References

[1.] Number Cruncher Statistical System (NCSS software, 2004)
[2.] K. M. Satter, "Competitive Usability Studies of Virtual Environments for Shipbuilding." PhD Dissertation, Engineering Management, University of New Orleans, 2005.
[3.] Syed Adeel Ahmed, "Usability Studies with Virtual and Traditional Computer Aided Design Environments." PhD Dissertation, Engineering Management, University of New Orleans, 2006.
[4.] Kurt Satter and Alley Butler, "Finding the Value of Immersive, Virtual Environments Using Competitive Usability Analysis", Transactions of the ASME, Journal of Computing and Information Science in Engineering, June 2012, Vol, 12.
[5.] **Dr. Syed Adeel Ahmed,** & Dr. Kurt M. Satter, (2013), Usability Studies in Virtual and Traditional Computer Aided Design Environments for Benchmark 1 (Navigation), International Journal of information Management & Information Systems, Vol.17, number 4. http://www.cluteinstitute.com/ojs/index.php/IJMIS/article/view/8096